

Pilot Review of MEDIN Revised Data Guidelines:

High level report in support of accompanying guidelines completed and summary overarching feedback on each guideline

Contents

1. INTRODUCTION	2
1.1. Aims	2
1.2. Stages complete	2
2. PERMISSION FROM DATA OWNERS (STAGE 1).....	3
3. EVALUATION (STAGE 2)	3
3.1. Internal assessment.....	3
3.2. Data owner review	4
4. GENERAL FEEDBACK (STAGE 3).....	5
4.1. Spreadsheet structure and data input process	5
4.2. Wider Implications and Comments	6

Report completed for the Marine Environment Data and Information Network (MEDIN) by Marine Planning Consultants (part of the Gardline Group)

Report compiled by Caroline Chambers with contributions from Paul Salmon (MPC), Dan Brutto (MESL) Rebecca Grice (Geosurveys), Alison Collier (Geosurveys), Mark Barham (GEL)

Version Control

<i>Draft Final Report</i>	<i>15/04/2014</i>	<i>MEDINPilot_Report_Submitted150414.docx</i>
<i>Final Report V1</i>	<i>22/04/2014</i>	<i>MEDINPilot_Report_Submitted220414.docx</i>
<i>Final Report V2</i>	<i>24/06/14</i>	<i>MEDINPilot_Report_Submitted240614.docx</i>

1. Introduction

1.1. Aims

The aim of the project was to pilot three selected MEDIN data guidelines:

- Species and benthos data by trawl or dredge
- Moored oceanographic instrument data
- Geophysical multi-channel seismic

The aim was to assess these through datasets collected by the Gardline Group where permission was provided by the client. The pilot was carried out in three stages as follows:

- **STAGE 1** – Source data and developer’s permission for three pilot datasets
- **STAGE 2** – Enter data and assessment of each field into spreadsheet template and where possible gain developers input
- **STAGE 3** – Assess process against common criteria and consolidate assessments into a brief report

The purpose of this document is to summarise feedback on the three guidelines. However also see the accompanying spreadsheet data guidelines completed.

1.2. Stages complete

Table 1 below shows the level of completion of each guideline.

Table 1. Level of completion of reach guideline

Guideline	Gardline Company	Stage 1: Permission	Stage 2: Evaluation	Stage 3: Combined assessment
Species and benthos data by trawl or dredge	Marine Ecological Surveys (MESL)	YES	YES But no contribution from data owner	YES
Moored oceanographic instrument data	MetOcean, Gardline Environmental Limited (GEL)	YES	YES But no contribution from data owner	YES
Geophysical multi-channel seismic	Geosurveys	YES	YES Metadata complete but data unable to be retrieved from archive and no contribution from data owner	YES

2. Permission from data owners (Stage 1)

Permission to use data was obtained from data owners for each of the three intended guidelines. However in all cases, permission was never obtained with the first enquiry and required a number of further enquiries until gained. The permission to use data generally relied on good client relationships where trust was built in former projects. However where clients did not approve of data being used, on the whole this was a time issue of gaining permission through the ranks of the companies within the timescales. For example one company was entering a busy period and would require approval from all four parent companies before they could release the data and this would not have been possible within the timeframes. Where successful in obtaining permission, time constraints similarly affected clients' feedback on our evaluations (see Stage 2).

Overall, obtaining permission and involvement from data owners required a certain amount of chasing up, encouragement and explanation of the intended purpose to get clients on side of the pilot project. Data owners were willing to participate but restricted by certain pressures and more pressing company needs.

3. Evaluation (Stage 2)

3.1. Internal assessment

See accompanying spreadsheets

- MEDIN_Seismic_1_4_Submitted150414
- MEDIN_Trawl_Dredge_4_0_Submitted150414
- MEDIN_Moored_Ocean_4_1_Submitted220414

These spreadsheets include completion of the data in the "Form" worksheets with the actual data plus the "Guidance" worksheets to respond to questions on completing the data. These include comments particular to the guideline, a summary of which is provided in **Section 4**.

Level of Completion

For seismic all "Form" worksheets were completed except for:

- Station Form: site centre only available as data cannot be retrieved from archive
- Single Channel Seismic Form: data cannot be retrieved from archive

It was the intention to use data that was available however due to the protracted communications on gaining permission, this was not possible, hence use of a dataset that was archived.

For benthic and epibenthic trawl all "Form" worksheets were completed expect for:

- Species Form: provided in Species Form Matrix Format worksheet provided instead
- Biotope Form: not provided in project that data is associated with
- Attribute Form: not provided in project that data is associated with

- Geological Data: not provided in project that data is associated with

For moored oceanographic data, all “Form” worksheets were completed.

Justification of Responses

The spreadsheets include completion of the “Guidance” worksheets with some fields added by MPC in blue text. These ask the question (input by MPC) “Did we complete this field?” to choose from a list of:

- YES - READILY AVAILABLE
- YES - BUT QUICK PROCESSING REQD
- NO - PROCESSING TOO EXTENSIVE
- NO - CANNOT BE WITHELD
- NO - NOT RELEVANT
- NO - NOT AVAILABLE

These responses are supported where relevant by “Comments to justify any 'No's”.

3.2. Data owner review

In addition the “Guidance” worksheets ask questions relating to the data owners views: “In theory, could this be made available on the public MEDIN Discovery Portal? “; and “In theory, could this be made available to a DAC (does not need to be public) - ?”. For those where permission was obtained, none of the data owners were able to provide comment. The reason for this was:

- Agreement initially given to contribute but despite multiple follow up contact (email, phone), developer too busy to do so. Their involvement in specific large scale work effecting all their marine developments is particularly high profile at the moment and takes precedence, understandably.
- Agreement initially given but developer was contacted towards the end of the pilot after others had fallen through and so time constraints. Comment was made on the actual data fields etc in liaison with their in house specialists but not on making metadata / data available. Though efforts were made to get their time for comment on the phone, responses were not made in time and pending as of date of this report. This was also due to the request for comment going through the company to reach the right person

One developer commented that they were sceptical about MEDIN compared to the System of Industry Metocean data for the Offshore and Research Communities (SIMORC)¹. They did comment that all databases should be consistent in terms of standards and link to each other; and that BODC seem to be member of both projects which would help guarantee this consistency.

¹ <http://www.simorc.org/>

4. General feedback (Stage 3)

A summary of the review is provided below. This summarises all feedback including that in **Table 2**, from the three individuals completing each guideline spreadsheet.

4.1. Spreadsheet structure and data input process

The spreadsheet structure seemed to present differing responses from each data type / person completing, though all were generally happy with or understanding of the structure.

Metadata

- Whilst clients were not available to comment, for the seismic data guideline, it was noted that in future we could clarify whether data / metadata could be made available to DACs / the MEDIN Discovery Portal as part of our contract specification.
- For each metadata field the user should make a response. Therefore if they have no valid entry, then they should enter “N/A” (not available) or “Cannot be made available”. This would avoid mis-interpretation of the feedback.

Data Fields - General

Overall, most fields were either readily available or not relevant. Very few fields required substantial time to source (see individual spreadsheet responses). The following are recommended for completion of fields:

- For the project start and end date it should be made absolutely clear that this should not equal the survey dates, but instead be more specific e.g. contract start until sign off when reporting complete.
- Put a guide on the length of the abstract to avoid one line responses due to lack of understanding the importance of this field, this would benefit from an example entered (currently absent) and number of words as a guide.

Data Fields – Database Structure

There is a small amount of mis-communication around the old database structure that is embedded into the new structure. The recommendations below address the main issues:

- Some of the fields could be automated where they are carried over into subsequent worksheets, e.g. methodID.
- If a field is mandatory but may not have been created for the project itself, e.g. methodID, stationID, then provide instructions on how to create a new ID so that users do not enter nothing.

Worksheet Labelling

To address some of the comments that the guidelines are long and complex, it is recommended that the worksheet labels are made much shorter so all of them can be seen at once, e.g. by an ID instead

of text label. For example “General Metadata Guidance, General Metadata Form, Detailed Metadata Guidance, Detailed Metadata Form” would become “1a, 1b, 2a, 2c”. These numbers could then be identified through a separate worksheet “Key”. This would help give the notion that they are manageable and help attract greater take-up of the guideline. This is a quick solution to help the guidelines seem less confusing to the new onlooker.

Streamlining

Whilst it is recognised that the MEDIN guidelines have been made simpler, feedback from staff not previously associated with MEDIN still responded that a more streamlined approach would be of value. This was noted in the context of reducing the volume of information requested (number of fields and worksheets). This response may have been more relevant to those guidelines that required a greater number of worksheets to be completed; and where data is related to a single station. However it is important to note these comments as any negative view of the guidelines will detract from their uptake generally.

Using Guidelines in the Field

In all cases the guidelines were considered fit for purpose in the field. However for moored oceanographic instrumentation, it was noted they would not replace existing deployment / recovery documentation or calibration sheets.

Time to Complete

It was estimated that an additional 1-3.5 hours would be required per dataset, resulting in fees of £40-250 to the contractor for completing the guideline. However as noted for the moored oceanographic data guideline, additional instruments and stations would significantly increase this time, e.g. up to a week. This has significant implications and it is recommended that multi-disciplinary surveys in oceanography are reviewed further, working together with experts in the field.

Consistency with GIS MEDIN Maestro

The fields are closely aligned to but not identical to those in the MEDIN Maestro metadata fields. It would help a great deal if these did align exactly to allow completion of information on the spreadsheet to be transferred over to GIS. This allows review in spreadsheet form for non-GIS managers, and quicker review generally even for GIS managers, as well as providing one central source for the information which avoids any variation in responses.

4.2. Wider Implications and Comments

The final point to make on the data guidelines relates to the feedback on the moored oceanographic guideline. In this field, other standards already exist, e.g. SIMROC, and it is recommended that these are brought into MEDIN as closely as possible to reduce time spent and increase take up.

MPC have previously contributed to the review of MEDIN carried out by Defra, some examples of which are provided below:

- Defra recommendation to provide direct access to data from the MEDIN Data Discovery Portal: we advised that this will only be reasonably obtainable for the public sector data, though recommended.
- Defra recommendation to improve MEDIN Data Discovery Portal: we advised that the area of interest could be drawn as an area, not a rectangle; and that feedback from irregular users would be useful also as the reason for irregular use are likely to be interface related. A possible improved use would be to integrate the discovery boundaries into a web feature service (or similar) to integrate into GIS systems such ArcGIS, MapInfo etc. This way people could take advantage of the discovery areas as part of their normal process.
- Defra recommendation to agree on the main end users/end products of MEDIN: This needs to be done carefully and should be undertaken after a greater marketing of MEDIN itself. Otherwise potential end users could be missed if they were to undertake this on the current user base as there is probably a diminished number of commercial users presently.

Table 2. Overarching feedback on completing the guidelines

	Seismic	Benthic and Epibenthic Trawl	Moored Oceanography
<i>Sharing of metadata and data</i>			
Completed By	Rebecca Grice (Gardline Geosurvey Ltd)	Daniel Brutto (Marine Ecological Surveys Ltd)	Mark Barham (Gardline Environmental Ltd)
Guideline Name	Geophysical multi-channel seismic MEDIN_Seismic_1_4	Species and benthos data by trawl or dredge MEDIN_Trawl_Dredge_4_0	Moored Instrument DataMEDIN_Moored_Ocean_4_1_MPC_V0 1
Can metadata be made available on the MEDIN Discovery Portal?	Data owner unavailable to comment. In future we could clarify this as part of our contract specification.	Data owner unavailable to comment.	Data owner unavailable to comment.
Can data be submitted to a Data Archiving Centre?	Data owner unavailable to comment. In future we could clarify this as part of our contract specification.	Data owner unavailable to comment.	Data owner unavailable to comment.
Does it require a licence?	Data owner unavailable to comment. In future we could clarify this as part of our contract specification.	Data owner unavailable to comment.	Data owner unavailable to comment.
Would they benefit from a provision of licence that can be tailored according to their needs?	Data owner unavailable to comment. We feel that a standardised licence could be evolved for this routine data.	Data owner unavailable to comment.	Data owner unavailable to comment.
Does it require payment from the data licensee?	Data owner unavailable to comment. Considered unlikely.	Data owner unavailable to comment.	Data owner unavailable to comment.
Would they benefit from a Memorandum of Understanding?	Data owner unavailable to comment. Probably unnecessary.	Data owner unavailable to comment.	Data owner unavailable to comment.

	Seismic	Benthic and Epibenthic Trawl	Moored Oceanography
<i>Spreadsheet structure</i>			
Is the spreadsheet structure and format fit for purpose?	It takes a short time to become familiar but otherwise it seems fit for purpose. The faded pastel colours on the original open up as fairly garish colours on different versions of Excel.	The spreadsheets are thorough and comprehensive and succeed in collecting a wealth of information about the data with which the spreadsheets are concerned. However, the numbers of different fields and worksheets required for completion are slightly concerning. The sheets themselves essentially collect and present information that are typically conveyed within two tables within our reports (covering metadata/fieldnotes and the data itself). It might be possible to streamline the volume of fields and tables required for completion. This would not only serve to decrease the duration required for completing this exercise, but would also probably increase uptake in participation in the system. It's important to recognise that the commercial industrial operators or their consultant who are responsible for completing these documents are paid for their contributions and if the process appears to be complicated and time consuming (even if it isn't) then it will discourage participation as people will seek to focus on pressing priorities relating to their day jobs. Simplification and streamlining are recommended as the biggest improvement that could be made to the process as it stands as this should encourage uptake and participation across the marine science community.	Some duplication in fields across the various tabs which could be automated.
Is the documentation fit for purpose?	It takes a short time to become familiar but otherwise it seems fit for purpose.	The supporting documents were comprehensive and assisted with the process of the completion of the metadata and provision of data. However, if steps could be taken to streamline the submission of this material, then the guidance documents too could be simplified and shortened.	A little cumbersome for small, single station measurement programs other than where incorporated into existing reporting. Acknowledged that detailed requested would be necessary for larger, prolonged oceanographic research cruises with multiple hydrographic stations and numerous pieces of instrumentation.
Do you have any comment with regards to other data guidelines?	No.	No.	Comment from data owner would be useful here

	Seismic	Benthic and Epibenthic Trawl	Moored Oceanography
<i>Data input process</i>			
Would it be possible to use this format in the field / during direct data collection / processing?	Yes, if it became standard, contractors would accommodate it within their workflow.	Using this in the field would be too time consuming and tricky, but potentially it could be incorporated into the project completion stages. As stated above, streamlining the process/required content to perhaps bring it more into line with standard (after a fashion) report appendices would probably encourage uptake and participation through promoting ease of use. A data owners' consultants would be the most likely compiler of data rather than the data owner themselves in most cases.	In field = no; during data collection = yes where tables can be incorporated into cruise reports ; during processing = yes. Some automation in duplicate fields would be useful. These spreadsheets would not replace existing deployment / recovery documentation or calibration sheets.
If no then what how much time does it take to translate it across post survey?	N/A	N/A	Single station with a single instrument would take an hour, additional instruments and stations would drastically increase this time. As indicated above some streamlining would alleviate this issue. E.g. 1-2 weeks linking existing reporting documentation into these sheets and/or modify existing documentation to be more MEDIN friendly.
What would the costs be to adopt these guidelines once 'business as usual'?	Negligible after setup.	There is a potential addition of 1-3.5 hours to each project to complete the required forms which would equate to approximately £40-250 in terms of opportunity costs if a consultancy/survey contractor was to complete the forms.	Not a simple process and would need to be done for documentation on each instrument
Do you have any comment with regards to other data guidelines?	No	No	No